



Contents list available at JMCS

Journal of Mathematics and Computer Science

Journal Homepage: www.tjmcs.com



Testing Adverse Selection Using Frank Copula Approach in Iran Insurance Markets

Hadi Safari Katesari^{1,*}, Behrouz Fathi Vajargah²

¹Department of Statistics, Shahid Beheshti University, Iran, ²Department of Statistics, University of Guilan, Iran

*ssafari.hadi@gmail.com

²fathi@guilan.ac.ir

Article history:

Received November 2014

Accepted February 2015

Available online February 2015

Abstract

Existence of adverse selection in insurance markets could have irreversible effects on enterprise decision-making process and obligations of insurance companies. In this article, testing adverse selection is done by jointly modeling the coverage selection and accidents frequency using Frank's copula, where the dependence parameter states the existence of relationship between coverage selection and the frequency of accidents. Our margins are modeled by ordered logistic regression model for the coverage selection and negative binomial regression model for the accidents frequency. The copula model is calibrated using 59,547 one-year cross-sectional cases of collision insurance coverage of Iran Insurance co. The results indicate a significant positive coverage selection-accidents frequency relationship.

Keywords: Adverse Selection, Copula, Logistic Regression, Collision Insurance.

1. Introduction

Ever since the seminal work of Shi and Valdez [6], statistician theorists have devoted much effort to using Archimedean copula models for research on asymmetric information. This paper searches on contribute to examin the application of these models on collision insurance data. An ordinary approach to testing adverse selection is correlation testing between contract selection and accidents frequency that we conduct this work with using of copula. In this article we follow Shi and Valdez [6] approach and we use a big sample data set that we caught from an insurer operating in the collision insurance market in Iran where contained 59,547 contracts. We use collision insurance for modeling and rank them in three level. Like the Shi and Valdez [6], we set accidents frequency in levels 0, 1, 2, 3 and 4 and for marginals distributions. Then we use ordered regression model and negative binomial. For testing adverse selection we employ Frank copula that it can show both negative and possitive association between margins. With fitting to data, our result show possitive association between policy selection and accident frequency.

These results calculated with dependence parameter $\theta=1.3$ and for testing our result we apply Clayton and Gumble copula. For the rest of the article, we have organized it as follows. Section 2 introduces the bivariate copula model for testing adverse selection. Section 3 describes the empirical data and calibrates the model using this data. In addition, the section discusses procedures used to examine model goodness-of-fit and its implications. A robust test of model implications is provided in this section. Section 4 concludes the paper with a discussion of additional further work.

2. Prior works on adverse selection in insurance market

Akerlof [1] demonstrated the problems arising from information asymmetry based on the “used car” market and referred to defective used cars as “lemons”. Recent empirical studies in the competitive automobile insurance market show no signs of adverse selection in these markets. To take some of the outstanding studies, Chiappori and Salani’e [2] find no systematic relationship between risk and coverage in the French automobile insurance market. These studies use data from the competitive insurance market and their empirical results suggest that insurers are successfully managing adverse selection or moral hazard, at least in the competitive insurance market where they can freely set premiums. For more information see Saito [5].

3. Model specification to testing adverse selection

For the first, we introduce the nature of collision insurance in Iran insurance co. With purchase a policy from Iran insurance co everybody coverages himself/herself from these main and overall claims: overall accident, overall theft and overall fire, as well as everybody can purchase one or more plethora coverage in this list: 1. Damaged caused by flood, earthquake and hurricanes, 2. Broken glass alone causes other than the main claims, 3. Instant stolen vehicle parts and accessories, 4. Damage caused by spills or splashes of paint, acid and chemicals, 5. Compensation by not using the vehicle in repair period, 6. Slippage (only in minor damage).

For the same argument then, we consider a multinomial measure and examine three types of coverage, ordinaly ranking them from lowest to highest: first level: overall coverage of collision insurance, second level: overall coverage of collision insurance in addition one or two plethora coverage(s) in the list, third level (comprehensive): overall coverage of collision insurance in addition three or four or five or all of plethora coverage(s) in the list. With a cross-sectional set of observations, we begin by letting y_{i1} and y_{i2} indicate the choice of coverage and accidents frequency, respectively, for policyholder i . Here, y_{i1} , with possible values of 1, 2, or 3, represents the choice of first level (overall), second level, and third level (comprehensive) coverages, respectively. Note that y_{i1} and y_{i2} are the observed variables whose values will be determined according to the corresponding latent variables defined by y_{i1}^* and y_{i2}^* , respectively. One could view y_{i1}^* as the policyholder’s preferred policy coverage and y_{i2}^* as the inherent risk level of the policyholder. We choose to model the observable variables and y_{i2} with a parametric copula to be denoted by $C(\cdot, \cdot)$. Then the joint probability mass function of y_{i1} and y_{i2} could be expressed as:

$$f_i(y_{i1}, y_{i2}) = C(F_{i1}(y_{i1}), F_{i2}(y_{i2})) - C(F_{i1}(y_{i1}-1), F_{i2}(y_{i2})) - C(F_{i1}(y_{i1}), F_{i2}(y_{i2}-1)) + C(F_{i1}(y_{i1}-1), F_{i2}(y_{i2}-1)) \quad (1)$$

where F_{i1} and F_{i2} are the cumulative distribution functions of y_{i1} and y_{i2} , respectively. Due to the parametric feature of the copula model, one needs specifications of the distribution function F_{i1} and F_{i2} for model identification. The coverage choice is measured on an ordinal scale. Thus, an ordered

multinomial model is used to describe the relationship between the response y_{i1} and the latent variable y_{i1}^* :

$$y_{i1} = \begin{cases} 1 & \text{if } y_{i1}^* \leq \alpha_1 \\ 2 & \text{if } \alpha_1 \leq y_{i1}^* \leq \alpha_2 \\ 3 & \text{if } y_{i1}^* > \alpha_2 \end{cases} \quad (2)$$

where α_1 and α_2 are unknown thresholds to be additionally estimated. We considered an ordered logistic regression model in the estimation. Henceforth, we use

$$F_{i1}(y_{i1}) = P(Y_{i1} \leq y_{i1}) = \begin{cases} \frac{1}{1 + \exp(-(\alpha_1 - x_i \beta))}, & y_{i1} = 1 \\ \frac{1}{1 + \exp(-(\alpha_2 - x_i \beta))}, & y_{i1} = 2 \\ 1, & y_{i1} = 3 \end{cases} \quad (3)$$

accidents frequency y_{i2} is specified using a negative binomial regression model. More specifically, its probability mass function is expressed as:

$$f_{i2}(y_{i2}) = P(Y_{i2} = y_{i2}) = \binom{\psi + y_{i2} - 1}{y_{i2}} \left(\frac{\psi}{\psi + \lambda_i} \right)^\psi \left(\frac{\lambda_i}{\psi + \lambda_i} \right)^{y_{i2}} \quad (4)$$

with ψ the dispersion parameter for policyholder i . The model specified in this section by its nature is fully parametric and can therefore be easily estimated using likelihood-based methods. To accommodate the fact that the choice of coverage and the frequency of accidents could possibly be either positively or negatively associated, we consider the Frank copula which permits such flexibility:

$$C(u_1, u_2; \theta) = -\frac{1}{\theta} \log \left[1 + \frac{(e^{-\theta u_1} - 1)(e^{-\theta u_2} - 1)}{e^{-\theta} - 1} \right], \quad (5)$$

where θ is the dependence parameter that captures the association between the two responses. The flexibility of allowing for either direction of association has been one of the primary reason for its popularity in applications in insurance, finance and medical statistics. Additional statistical properties of the Frank's family of copulas in (5) have been explored in Genest [3] and Nelsen [4].

4. Calibrating the model

Data used to calibrate the model specified in Section 3 was drawn from a portfolio of automobile insurance policies of a major insurer in Iran. In particular, we use the observations in calendar year 2007-2008 for this insurer where we have a total of 59,547 policies that were recorded in the collision insurance portfolio. Similar to several jurisdictions worldwide, Iran requires drivers to have, at the minimum, a third party liability coverage to be able to drive a vehicle on the road, and at the same time, drivers have the liberty to choose beyond this minimum level of coverage. Our data set indeed comes from a subsample of the Iran insurance co portfolio. Table 1 provides a summary of the frequency statistics for our two primary variables of interest.

Table1. Number and percentage of policy choice and reported accidents

Policy Choice	1	2	3	Total Number	Percent
0	30176	20033	4879	55088	92.51
1	405	1497	2130	4032	6.77
2	39	161	192	392	0.66
3	2	11	21	34	0.06
4	0	1	0	1	0.00
Total Number	30622	21703	7222	59547	
Percent	51.42	36.45	12.13		100

4.1 Estimation results and discussion

The resulting (maximum likelihood) estimates for the copula model are presented in table 2. In examining the effect of vehicle characteristics, we find that the age of the vehicle exhibits significant effect on both policy choice and accident occurrence. Another explanatory variable that is worth making an observation is the NCD (No Claim Discount) factor. First, there is a significant effect of NCD on policy choice in the sense that a policyholder with a high NCD tends to purchase better insurance coverage on its vehicle. Consistently, a driver with a lower NCD tends to have more accidents.

Table 2. Estimates of Frank copula model for all reported accidents

Choice - Cumulative Logit	Estimate	StdErr	Risk - Negative Binomial	Estimate	StdErr
Choice- α_1	-0.9032	0.0131			
Choice- α_2	0.3769	0.0126	Risk-intercept	-2.1588	0.0194
Choice-sex (woman)	0.4942	0.0195	Risk -sex (woman)	0.0189	0.7469
Choice-vehicle age	-0.0415	0.0034	Risk -vehicle age	0.0172	0.8194
Choice-(NCD=2)	0.0673	0.0199	Risk -(NCD=2)	0.4437	0.0390
Choice-(NCD=3)	0.2300	0.0191	Risk -(NCD=3)	-0.7787	0.0414
Choice-(NCD=4)	0.1632	0.0277	Risk -(NCD=4)	-1.3940	0.0828
Choice-vehicle application (2)	-1.5683	0.1907	Risk -vehicle application (2)	0.4333	0.3250
Choice-vehicle type (2)	-1.3206	0.6930	Risk -vehicle type (2)	-0.7912	0.1779
			Dispersion	0.8326	0.0739
Dependence parameter θ	1.3023	0.0573			
-2Loglikelihood	160170				

4.2 Quality of fit tests

Goodness-of-fit tests are performed for the marginals as well as for the copula. For marginal distributions, we exhibit the observed and fitted frequencies for both the policy choice and accidents frequency in table 3. The consistency between the actual and fitted frequencies suggests very satisfactory fit for both marginals. According to table 2, the estimation of dependence parameter in the Frank copula is roughly 1.3, which translates to a Spearman’s rho coefficient of roughly 22 percent. This provides an evidence of the positive association between the policy choice and level of risk of the policyholder. We find that drivers with better coverage tend to be more prone to make claims. This observation would be explained by the presence of adverse selection. A final examination of the copula model is a robustness test. We re-

calibrated the copula model under two other customarily used Archimedean-type copulas, the Gumbel copula and the Clayton copula. The estimated positive relationship that we already observed based on the Frank copula between policy choice and risk in Section 3, suggests both copulas are eligible to test for possible robustness. The dependence parameter of the Gumbel copula is 1.1 that translates to a Spearman’s rho of 0.15. The dependence parameter of the Clayton copula is 0.26 that corresponds to a Spearman’s rho of 0.20. Both models suggest a positive association between the policy selection and risk level of the policyholder. The results indicate a significant positive coverage selection-accidents frequency relationship.

Table 3. Goodness-of-fit tests of the marginals

	Choice			Risk	
Value	Observed	Fitted	Value	Observed	Fitted
1	51.42	52.32	0	92.51	92.50
2	36.45	34.29	1	6.77	6.80
3	12.13	13.39	2	0.66	0.62
			3	0.06	0.06
			4	0	0.01

5. Conclusion

In this paper we used a bivariate copula regression method to jointly examine the policyholder’s coverage choice and the level of risk. To test for the presence of adverse selection, the policyholder’s coverage selection was measured by an ordinal categorical variable and the degree of risk was approximated by an ex post risk measure, the number of times a policyholder has claimed in a calendar year. To calibrate the copula model, we used a cross-sectional empirical observation of an insurance portfolio from a major automobile insurer in Iran. After controlling for the risk factors (policyholder and vehicle characteristics) observed by the insurer, we found evidence of a strong positive coverage-risk association, which suggests the possible existence of private information by the policyholders.

References

- [1] G.A. Akerlof, “*The market for ‘lemons’: quality uncertainty and the market mechanism*”, Quarterly Journal of Economics, 84 (1970) 488-500.
- [2] P-A. Chiappori, B. Salanié, “*Testing for asymmetric information in insurance market*”, Political Economy. 108 (2000) 56-78.
- [3] C. Genest, “*Frank’s family of bivariate distributions*”, Biometrika. 73 (1987) 549-555.
- [4] R.B. Nelsen, “*An introduction to Copulas, Portland*”, Springer, New York, (2006).
- [5] K. Saito, “*Testing for asymmetric information in the automobile insurance market under rate regulation*”, The Journal of Risk and Insurance. 73 (2006) 335-356.
- [6] P. Shi, E.A. Valdez, “*A Copula approach to test asymmetric information with applications to predictive modeling, Insurance*” Mathematics and Economics. 49 (2011) 226-239.